Characterization of approximation schemes satisfying Shapiro's Theorem

J. M. Almira

Abstract

In this paper we characterize the approximation schemes that satisfy Shapiro's theorem and we use this result for several classical approximation processes. In particular, we study approximation of operators by finite rank operators and n-term approximation for several dictionaries and norms. Moreover, we compare our main theorem with a classical result by Yu. Brundyi and we show two examples of approximation schemes that do not satisfy Shapiro's theorem.

1 Introduction and motivation

One of the most important results in the constructive theory of functions is the so called Bernstein's lethargy theorem, which claims that if $X_0 \subset X_1 \subset X_2 \subset \cdots \subset X_n \subset \cdots \subset X$ is an ascending chain of finite dimensional vector subspaces of a Banach space X, being all strict inclusions, and $\{\varepsilon_n\} \searrow 0^+$ is a non-increasing sequence of positive real numbers that converges to zero, then there exists an element $x \in X$ such that $E(x, X_n) = \inf_{y \in X_n} \|x - y\| = \varepsilon_n$ for all $n \in \mathbb{N}$. This result was first obtained in 1938 by S.N. Bernstein [2] for $X = C[0,1]$ and $X_n = \Pi_n$, the vector space of real polynomials of degree $\leq n$. Later on, the theorem was generalized firstly to the result we have already stated and then to several technical results on quite general abstract spaces. Indeed there are essentially two ways in which a generalization of Bernstein's lethargy theorem was studied. The first one was to look for a result valid for arbitrary chains of (possibly infinite dimensional) closed subspaces of the Banach space X. In this direction the best results that exist were proved by Tjuriemskih [18] and Nikolskii [10, 11] and state that a sufficient condition for the validity of such a lethargy theorem is that $X = H$ is a Hilbert space and that a necessary condition is that $X = X^{**}$ (i.e., X must be a reflexive space), respectively. Another focus for a generalization of the lethargy theorem that has been deeply studied is to maintain the finite-dimensional chains of subspaces of X but looking for more general spaces X than those given by the Banach setting, where a distance function is well defined and a Lethargy theorem holds true. In this direction the work by G. Lewicki [7, 8] is, as far as we know, the best representative of successful results. In particular, he obtained several results in the context of SF-spaces. Finally, a mixture of both attempts of generalization has been made by Micherda [9]. It is also interesting to recall that, thanks to the work by Plesniak [14], in the seventies the lethargy theorem became a very useful tool for the theory of quasianalytic functions of several complex variables.

In 1964 H.S. Shapiro [15] observed that an easy consequence of Bernstein's theorem is the fact that for any non-increasing sequence $\{\varepsilon_n\} \searrow 0^+$ there exists an $x \in X$ such that $E(x, X_n) \neq \mathbf{O}(\varepsilon_n)$,

and proved the corollary by an elementary argument that uses Riesz's Lemma and the Baire category theorem. Furthermore, Shapiro's proof did not need any hypothesis on the dimension of X_n. He just assumed that X_n is a closed subspace of X for all n. By this way, his result was transformed from a simple corollary of Bernstein's lethargy theorem into a new interesting non-trivial result in approximation theory. He also proved an analogous result for generalized rational approximation.

However everybody knows that approximation by linear subspaces of a Banach space is a very restrictive process of approximation. There are many other choices of approximation processes such as rational approximation, approximation by splines with of without free knots, n-term approximation with dictionaries of different kinds, wavelets and approximation of operators by operators of finite rank, just to mention a few of them. So, it seems an interesting question to know in each case if Bernstein's result or Shapiro's result holds true. In this precise sense and with respect to Bernstein's lethargy theorem, the most general result that exists was proved by Yu. Brudnyi [3]. It claims that if X is a Banach space, $\{0\} = A_0 \subset A_1 \subset \cdots \subset A_n \subset \cdots X$ is an infinite chain of subsets of X and the A_n satisfy the conditions

- $A_n + A_m \subset A_{n+m}$ For all $n, m \in \mathbb{N}$.
- $\lambda A_n \subset A_n$ for all $n \in \mathbb{N}$ and all scalar λ.
- $\bigcup_{n\in\mathbb{N}} A_n$ is a dense subset of X

and

$$\gamma = \inf_{n\in\mathbb{N}} dist(A_{n+1} \cap S(X), A_n) > 0, \tag{1}$$

(where $S(X)$ denotes unit sphere of X and $dist(A, B) = \sup_{a\in A} d(a, B)$ with $d(a, B) = \inf_{b\in B} \|a - b\|_X$) then for every non-increasing convex sequence $\{\varepsilon_n\}_{n=0}^{\infty} \searrow 0^+$ there is some $x \in X$ such that $E(x, A_n) \geq \varepsilon_n$ for all $n \in \mathbb{N}$. Obviously this theorem is weaker than the lethargy theorem since it imposes an important restriction on the sequence $\{\varepsilon_n\} \searrow 0^+$ (been convex) and it also loses the equalities $E(x, A_n) = \varepsilon_n$ but, on the other hand, it is stronger than Shapiro's theorem since the inequalities $E(x, A_n) \geq \varepsilon_n$ are guaranteed for all $n \in \mathbb{N}$. Originally, this result was published in 1981 in Russian by Yu. Brudnyi and N. Ya. Krugljak (although the paternity belongs to Brudnyi) as part of their well known monograph on interpolation theory and only in 1991 the results was exposed to the Anglo-Saxon community, when the monograph was translated into English (see [3]).

In this paper we characterize the approximation schemes that satisfy Shapiro's theorem (see Definition 2 and Theorem 4) and we use this result for several classical approximation processes. In particular, we study approximation of operators by finite rank operators and n-term approximation for several dictionaries and norms. Moreover, we compare our main theorem with Brundyi's theorem and we show two examples of approximation schemes that do not satisfy Shapiro's theorem.

2 The main result

Before proving the main result of this paper, we introduce the general concept of approximation scheme and give a precise meaning to the phrase "to satisfy Shapiro's theorem" for approximation schemes. Moreover, we state and prove a technical lemma about sequences of real numbers.

Definition 1 *Let $(X, \|\cdot\|)$ be a quasi-Banach space and let $A_0 \subset A_1 \subset \cdots \subset A_n \subset \cdots X$ be an infinite chain of subsets of X, where all inclusions are assumed to be strict. We say that $(X, \{A_n\})$ is an approximation scheme whenever the following conditions hold true:*

(i) *There exists a map $K : \mathbb{N} \to \mathbb{N}$ such that $K(n) \geq n$ and $A_n + A_n \subseteq A_{K(n)}$ for all $n \in \mathbb{N}$.*

(ii) *$\lambda A_n \subset A_n$ for all $n \in \mathbb{N}$ and all scalar λ.*

(iii) *$\bigcup_{n\in\mathbb{N}} A_n$ is a dense subset of X*

Definition 2 *We say that $(X, \{A_n\})$ satisfies Shapiro's theorem if for all non-increasing sequence $(\varepsilon_n) \searrow 0^+$ there exists some $x \in X$ such that $E(x, A_n) \neq \mathbf{O}(\varepsilon_n)$.*

Lemma 3 *Let $h : \mathbb{N} \to \mathbb{N}$ be a map such that $h(n) \geq n$ for all n and let $\{\varepsilon_n\} \searrow 0^+$. Then there exists a sequence $\{\xi_n\} \searrow 0^+$ such that $\xi_n \geq \varepsilon_n$ and $\xi_n \leq 2\xi_{h(n)}$ for all n.*

PROOF. Firstly, we prove that there are sequences (a_n) such that $(a_n) \searrow 0^+$ and $a_n \leq 2a_{h(n)}$ for all n. To do this, we assume that h is strictly increasing and $h(1) > 1$ (otherwise, we set $h^*(n) = \max\{h(0), ..., h(n)\} + n$, and prove the result for h^*. This will be enough since $a_n \leq 2a_{h^*(n)}$ implies $2a_{h(n)} \geq 2a_{h^*(n)} \geq a_n$).

It follows from our hypothesis on h that $\lim_{s\to\infty} h^s(1) = \infty$, where $h^{s+1}(n) = h(h^s(n))$ for all $s, n \in \mathbb{N}$. Now, we set

$$a_n = \begin{cases} 1 & \text{if } n \in \{0,1\} \\ 1/2^s & \text{if } n \in \{h^s(1), h^s(1)+1, ..., h^{s+1}(1)-1\} \end{cases} \quad (s \in \mathbb{N})$$

It is clear that $\{a_n\} \searrow 0^+$. On the other hand, let $s = s(n)$ be such that $h^s(1) \leq n < h^{s+1}(1)$, then $h^{s+1}(1) \leq h(n) < h^{s+2}(1)$ and $a_n = 2a_{h(n)}$.

Set $b_n = \max\{a_n, \varepsilon_n\}$. It is clear that $b_n \geq \varepsilon_n$ for all n but it could happen that $\sup_{n\in\mathbb{N}}\{b_n/b_{h(n)}\} = \infty$. Now we set $\xi_0 = b_0$ and

$$\xi_{n+1} = \begin{cases} \xi_n & \text{if } \xi_n - b_{n+1} < a_n - a_{n+1} \\ \xi_n - (a_n - a_{n+1}) & \text{if } \xi_n - b_{n+1} \geq a_n - a_{n+1} \end{cases}, (n \in \mathbb{N})$$

Then

$$\xi_{h(n)} = \xi_{n+(h(n)-n)} \geq \xi_n - \sum_{k=n}^{h(n)-1} (a_k - a_{k+1}) = \xi_n - (a_n - a_{h(n)}) > 0$$

for all n. Hence

$$\frac{\xi_n}{\xi_{h(n)}} \leq \frac{\xi_n}{\xi_n - (a_n - a_{h(n)})} \leq \frac{a_n}{a_n - (a_n - a_{h(n)})} = \frac{a_n}{a_{h(n)}} \leq 2, (n \in \mathbb{N});$$

since $\xi_n \geq a_n$ and the function $f(x) = \frac{x}{x-\alpha}$ is decreasing on $(\alpha, +\infty)$. $\Box$

Theorem 4 *The following are equivalent claims:*

(a) *The approximation scheme $(X, \{A_n\})$ satisfies Shapiro's theorem.*

(b) *There exists a constant* $c > 0$ *and an infinite set* $\mathbb{N}_0 \subseteq \mathbb{N}$ *such that for all* $n \in \mathbb{N}_0$, *there exists some* $x_n \in X \setminus \overline{A_n}$ *which satisfies* $E(x_n, A_n) \leq cE(x_n, A_{K(n)})$.

PROOF. Let us prove that (b) implies (a). As a first step, we will prove the result under the additional hypothesis that ε_n satisfies the inequalities: $\varepsilon_n \leq 2\varepsilon_{K(n+1)-1}$ for all $n \in \mathbb{N}$. So, let us now assume that $E(x, A_n) = \mathbf{O}(\varepsilon_n)$ for all $x \in X$. Then $X = \bigcup_{m=1}^{\infty} \Gamma_m$, where $\Gamma_\alpha = \{x \in X : E(x, A_n) \leq \alpha\varepsilon_n, n = 0, 1, 2, \cdots\}$ for all $\alpha > 0$. The sets Γ_m are closed subsets of X, so that we can use Baire's lemma to claim that there exists some $m_0 \in \mathbb{N}$ such that Γ_{m_0} has non empty interior. This means that there exists a ball $B(x, r) \subset \Gamma_{m_0}$ with $r > 0$. Now, $E(-x, A_n) = E(x, A_n)$ for all n, so that $\Gamma_m = -\Gamma_m$ for all m. In particular, $-B(x, r) \subset \Gamma_{m_0}$. Let us now take $z = \lambda x + (1 - \lambda)y$ a convex linear combination of two elements $x, y \in \Gamma_{m_0}$. Then

$$\begin{aligned}
E(z, A_{K(n)}) &= \inf_{g \in A_{K(n)}} \|\lambda x + (1-\lambda)y - g\| \\
&\leq \inf_{a,b \in A_n} \|\lambda(x-a) + (1-\lambda)(y-b)\| \\
&\leq C_X[\inf_{a \in A_n} \|\lambda(x-a)\| + \inf_{b \in A_n} \|(1-\lambda)(y-b)\|] \\
&= \lambda C_X E(x, A_n) + (1-\lambda) C_X E(y, A_n) \leq m_0 C_X \varepsilon_n,
\end{aligned}$$

since $A_n + A_n \subseteq A_{K(n)}$ and $\alpha A_n \subseteq A_n$ for all scalar α. On the other hand, the condition imposed on the sequence $\{\varepsilon_n\}_{n=0}^{\infty}$ implies that

$$\begin{aligned}
\sup_{n \in \mathbb{N}} \frac{E(z, A_n)}{\varepsilon_n} &= \sup_{m \in \mathbb{N}} \max \left\{ \frac{E(z, A_n)}{\varepsilon_n} \right\}_{n=K(m)}^{K(m+1)-1} \\
&\leq \sup_{m \in \mathbb{N}} \max \left\{ E(z, A_{K(m)}) \frac{1}{\varepsilon_n} \right\}_{n=K(m)}^{K(m+1)-1} \\
&= \sup_{m \in \mathbb{N}} E(z, A_{K(m)}) \frac{1}{\varepsilon_{K(m+1)-1}} \\
&\leq \sup_{m \in \mathbb{N}} E(z, A_{K(m)}) \frac{2}{\varepsilon_m} \leq 2m_0 C_X.
\end{aligned}$$

Hence $z \in \Gamma_{2m_0C_X}$. It follows that for a certain positive radius $\rho > 0$, the ball $B_\rho = \{x \in X : \|x\| \leq \rho\}$ is a subset of $\Gamma_{2m_0C_X}$. Hence, for every $x \in X$ we have that $\frac{\rho}{\|x\|} x \in \Gamma_{2m_0C_X}$ and the inequality

$$E(x, A_n) \leq \frac{\|x\|}{\rho} 2m_0 C_X \varepsilon_n$$

holds true for all $x \in X$ and all $n \in \mathbb{N}$.

Take $n \in \mathbb{N}_0$ and let $a_n \in A_n$ be an element of the cone A_n verifying $\|x_n - a_n\| \leq 2E(x_n, A_n)$, where $\{x_k\}_{k \in \mathbb{N}_0}$ is the sequence of elements of X given by condition (b). Let us take $y_n = x_n - a_n$. Then

$$\|y_n - b_n\| = \|x_n - (a_n + b_n)\| \geq E(x_n, A_{K(n)}) \geq \frac{1}{c} E(x_n, A_n) \geq \frac{1}{2c} \|y_n\|$$

for all $b_n \in A_n$. Hence

$$\frac{1}{2c} \|y_n\| \leq E(y_n, A_n) \leq \frac{\|y_n\|}{\rho} 2m_0 C_X \varepsilon_n$$

for all $n \in \mathbb{N}_0$. Dividing by $\|y_n\|$ everywhere at the inequalities above, we get

$$\frac{1}{2c} \leq \frac{1}{\rho} 2m_0 C_X \varepsilon_n,$$

which is in contradiction with $\varepsilon_n \to 0$. This proves the result for sequences $\{\varepsilon_n\}_{n=0}^{\infty}$ verifying the inequalities $\varepsilon_n \leq 2\varepsilon_{K(n+1)-1}$, $n \in \mathbb{N}$.

Let us now assume that $\{\varepsilon_n\}_{n=0}^{\infty}$ is an arbitrary non-increasing sequence which converges to zero for n approaching infinity. It follows from the application of Lemma 3 for the sequence $\{\varepsilon_n\}_{n=0}^{\infty}$ and the map $h(n) = K(n+1) - 1$, that there exists a sequence $\{\xi_n\}_{n=0}^{\infty}$ that satisfies the inequalities $\xi_n \leq 2\xi_{K(n+1)-1}$ and $\xi_n \geq \varepsilon_n$ for all $n \in \mathbb{N}$. This ends the proof of $(b) \Rightarrow (a)$ since for this new sequence we have already proved the existence of an element $x \in X$ such that $E(x, A_n) \neq \mathbf{O}(\xi_n)$, which implies $E(x, A_n) \neq \mathbf{O}(\varepsilon_n)$.

Now we prove that (a) implies (b). If $X = \cup_{n=0}^{\infty} \overline{A_n}$ then both (a) and (b) are false, since in such a case the sequences of errors $E(x, A_n)$ are stationary at zero. Hence we can assume that $X \neq \cup_{n=0}^{\infty} \overline{A_n}$ without loss of generality. If (b) is false, the sequence $\{c_n\}_{n=0}^{\infty} \subset [0, \infty)$ given by

$$c_n = \inf_{x \in X \setminus \overline{A_{K(n)}}} \frac{E(x, A_n)}{E(x, A_{K(n)})}$$

satisfies $\lim_{n\to\infty} c_n = \infty$, since it has no bounded subsequences. If we set $\varepsilon_k = 1/c_n$ for each $k \in [K(n), K(n+1))$, and we take $x \in X \setminus \cup_{n=0}^{\infty} \overline{A_n}$ then for each $k \in [K(n), K(n+1))$,

$$E(x, A_k) \leq E(x, A_{K(n)}) \leq \frac{1}{c_n} E(x, A_n) \leq \frac{1}{c_n} \|x\| = \varepsilon_k \|x\|$$

so that, $E(x, A_k) = \mathbf{O}(\varepsilon_k)$ and (a) is also false. This ends the proof. □

It follows from Theorem 2 that every linear approximation scheme (i.e. every approximation scheme verifying $K(n) = n$ for all n) satisfies Shapiro's theorem. In particular, this proves Shapiro's theorem for quasi-Banach spaces. Moreover, if X is a space of functions $f : [a, b] \to \mathbb{R}$ which contains a sequence of equioscilanting functions $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that f_{n_k} equioscillates n_k times and the cones A_n satisfy a Tchebychev's alternation principle (i.e., there exists a natural number $r(n)$ such that if $a \in A_n$ and $f - a$ equioscillates $r(n)$ times in $[a, b]$ then $E(f, A_n) = \|f - a\|_X$) then the approximation scheme $(X, \{A_n\})$ satisfies Shapiro's theorem, since $E(f_{n_k}, A_{K(n)}) = \|f_{n_k}\|_X = E(f_{n_k}, A_n)$ for $k = k(n)$ large enough. In particular, this result implies that, for the uniform norm, rational approximation and all kinds of spline approximation based on polynomials and rational functions satisfy Shapiro's theorem.

3 Approximation of operators $T : X \to X$ by operators of finite rank

Theorem 5 *Let us assume that there exists a sequence $\{P_n\}_{n\in\mathbb{N}_0}$ of linear projections $P_n : X \to X$ of finite rank, $\mathbf{rank}(P_n) = n$ for all $n \in \mathbb{N}_0$, such that $\sup_{n\in\mathbb{N}_0} \|P_n\| = C < \infty$. Then for all non-increasing sequence $\{\varepsilon_n\} \searrow 0^+$ there are approximable operators T such that $a_n(T) \neq \mathbf{O}(\varepsilon_n)$, where $a_n(T) = \inf_{\mathbf{rank}(R)<n} \|T - R\|$ denotes the n-th approximation number of the operator T.*

PROOF Let $H_n = P_n(X)$ be the range of P_n and define $Q_n : X \to H_n$ by $Q_n(x) = P_n(x)$. Let us denote by $i_n : H_n \to X$ the inclusion map. Then $1_{H_n} = Q_n P_n i_n$, $\|i_n\| = 1$ and $\|Q_n\| = \|P_n\| \leq C$. Hence

$$1 = a_n(1_{H_n}) \leq \|Q_n\| a_n(P_n) \|i_n\| \leq C a_n(P_n)$$

This obviously implies that

$$a_{[n/2]}(P_n) \leq \|P_n\| \leq C = C^2 \frac{1}{C} \leq C^2 a_n(P_n) \text{ for all } n \in \mathbb{N}_0$$

and the proof follows using Theorem 2 for the approximation scheme

$$(F(X,X) = \{T : X \to X : a_n(T) \searrow 0\}, \{\Sigma_n = \{R : X \to X : \mathbf{rank}(R) < n\}\}_{n=1}^{\infty}).$$

□

Corollary 6 *Let us assume that there exists a projection $P : X \to X$ such that the space $Y = P(X)$ has a Schauder basis. Then for all non-increasing sequence $\{\varepsilon_n\} \searrow 0^+$ there are approximable operators $T \in F(X,X)$ such that $a_n(T) \neq \mathbf{O}(\varepsilon_n)$. In particular, the same result holds true if X has a Schauder basis.*

PROOF Let $\{x_n\}_{n=1}^{\infty}$ be a Schauder basis of $Y = P(X)$ and let $U_n : Y \to Y$ denote the projection $U_n(x) = \sum_{k=1}^{n} a_i x_n$, where $x = \sum_{k=1}^{\infty} a_i x_n$. It is well known that $\sup_{n\in\mathbb{N}} \|U_n\| < \infty$. Hence $\sup_{n\in\mathbb{N}} \|iU_n Q\| < \infty$, where $i : Y \to X$ is the inclusion map and $Q : X \to Y$ is given by $Q(x) = P(x)$ for all $x \in X$. Hence we can use Theorem 3 with $P_n = iU_n Q$, $n \in \mathbb{N}$. □

4 n-term approximation

In this section we study Shapiro's theorem for n-term approximation. To do this, we need first to recall a few concepts and notations.

Let X be a Banach space. We say that $\mathcal{D} \subset X$ is a dictionary of X if $\mathbf{span}(\mathcal{D})$ is a dense subspace of X. In this case we define the approximation scheme $(X, \Sigma_n(\mathcal{D}))$, where

$$\Sigma_0(\mathcal{D}) = \{0\}; \ \ \Sigma_n(\mathcal{D}) = \bigcup_{\{\phi_{k_1},\cdots,\phi_{k_n}\}\subset\mathcal{D}} \mathbf{span}\{\phi_{k_1}, \phi_{k_2}, \cdots, \phi_{k_n}\} \ \ (n \geq 1). \tag{2}$$

and, associated to it, we study the errors of best n-term approximation:

$$\sigma_n(x, \mathcal{D}) = E(x, \Sigma_n(\mathcal{D})) = \inf_{z\in\Sigma_n(\mathcal{D})} \|x - z\|.$$

Obviously the properties of the sequence of errors $\sigma_n(x, \mathcal{D})$ strongly depend on the dictionary $\mathcal{D}$. For example, if $\overline{\mathcal{D}}^X = X$, then $\sigma_n(f, \mathcal{D}) = 0$ for all $n \geq 1$ and the dictionary is of no interest. On the other hand, a very reasonable choice of dictionary is $\mathcal{D} = \{\varphi_k\}_{k=1}^{\infty}$ a Schauder basis of X such that $\|\varphi_k\| = 1$ for $k = 1, 2, \cdots$ (we say that $\mathcal{D}$ is normalized). With this choice, any element $x \in X$ admits a unique representation of the form $x = \sum_{k=1}^{\infty} c_k(x)\varphi_k$. This allow us to introduce the

concept of greedy approximation. Concretely, for each $x \in X$ we define the set $D(x)$ of permutations $\rho : \mathbb{N} \to \mathbb{N}$ such that

$$|c_{\rho(j)}(x)| \geq |c_{\rho(j+1)}(x)|; j = 1, 2, \cdots$$

and, for each $\rho \in D(x)$ we define the n-th greedy approximation of x with respect to the basis $\mathcal{D}$ and the permutation ρ as

$$G_n(x, \mathcal{D}, \rho) = \sum_{k=1}^{n} c_{\rho(k)}(x)\varphi_{\rho(k)}.$$

We say that the basis $\mathcal{D}$ is greedy if there exists a constant $C = C(D, X)$ such that for every $x \in X$ there exists a permutation $\rho \in D(x)$ such that

$$\|x - G_n(x, \mathcal{D}, \rho)\| \leq C\sigma_n(x, \mathcal{D}). \tag{3}$$

This concept was introduced by Konyagin and Temlyakov [6] in 1999. In that paper they proved that for any greedy basis $\mathcal{D}$ of a Banach space X the inequality (3) holds true for all $\rho \in D(x)$. In other words, they proved that been a greedy basis is equivalent to say that

$$\delta_{X,\mathcal{D}}(n) = \sup_{x \in X \setminus \overline{\Sigma_n(\mathcal{D})}, \rho \in D(x)} \frac{\|x - G_n(x, \mathcal{D}, \rho)||}{\sigma_n(x, \mathcal{D})} = \mathbf{O}(1).$$

Moreover, they also got the following characterization of these bases:

Theorem 7 (Konyagin & Temlyakov) *Let $\mathcal{D} = \{\varphi_n\}_{n=1}^{\infty}$ be a normalized Schauder basis of the Banach space X. Then the following are equivalent claims:*

(a) *$\mathcal{D}$ is greedy.*

(b) *$\mathcal{D}$ is unconditional and democratic.*

Recall that a Schauder basis $\mathcal{D} = \{\varphi_n\}_{n=1}^{\infty}$ is unconditional if for every $x \in X$ the series $x = \sum_{k=1}^{\infty} c_k(x)\varphi_k$ is unconditionally convergent. On the other hand, the basis $\mathcal{D}$ is democratic whenever there is a constant $C > 0$ such that for every two finite subsets Λ, Λ^* of $\mathbb{N}$, if they have the same cardinality $|\Lambda| = |\Lambda^*|$, then

$$\left\| \sum_{k \in \Lambda} \varphi_k \right\| \leq C \left\| \sum_{k \in \Lambda^*} \varphi_k \right\|.$$

For example, any orthonormal basis of a separable Hilbert space is unconditional and democratic (hence greedy). Another example of greedy basis is the univariate Haar basis of $L^p(0, 1)$, which is given by $\mathcal{H}_p = \{h_k\}_{k \in \mathbb{N}}$, where

$$h_{2^j+t} = 2^{j/p} \left(\chi_{\Delta_{2^{j+1}+2t-1}} - \chi_{\Delta_{2^{j+1}+2t}} \right); 1 \leq t \leq 2^j, j \geq 1,$$

$\Delta_{2^j+t} = [2^{-j}(t-1), 2^{-j}t]$ for $1 \leq t \leq 2^j$, $j \geq 1$, and χ_Δ denotes the characteristic function associated to the interval Δ (see [17] for the proof that $\mathcal{H}_p$ is greedy in $L^p(0, 1)$). Moreover, in [17] it was also proved that every basis of $L^p(0, 1)$ which is L_p-equivalent to $\mathcal{H}_p$ is greedy. Here the L^p-equivalence of the basis $\{\varphi_k\}_{k=1}^{\infty}$ with $\mathcal{H}_p$ means that there are two positive constants C_1, C_2 such that for any finite set $\Lambda \subset \mathbb{N}$ and any coefficients $\{c_k\}_{k \in \Lambda}$ we have that

$$C_1 \| \sum_{k \in \Lambda} c_k\varphi_k\|_{L^p} \leq \| \sum_{k \in \Lambda} c_k h_k\|_{L^p} \leq C_2 \| \sum_{k \in \Lambda} c_k\varphi_k\|_{L^p}.$$

Theorem 8 *If H is a separable Hilbert space and $\mathcal{D} = \{\varphi_k\}_{k=1}^{\infty}$ is an orthonormal basis of H, the approximation scheme $(H, \{\Sigma_n(\mathcal{D})\}_{n=0}^{\infty})$ satisfies Shapiro's theorem. In particular, for $H = L^2(\mathbb{R}^d)$ and $\mathcal{D} = \{\varphi_{j,k} = |\det(A)|^{j/2}\phi(A^j x - k)\}_{j\in\mathbb{Z},k\in\mathbb{Z}^d}$ any wavelet basis of H (with dilation matrix A), the associated n-term approximation process satisfies Shapiro's theorem.*

PROOF We know that every orthonormal basis of H is greedy. In fact, in this case it is easy to check that, for all $x \in H$ and $\rho \in D(x)$,

$$\|x - G_n(x, \mathcal{D}, \rho)\| = \sigma_n(x, \mathcal{D}) = \sqrt{\sum_{k=n+1}^{\infty} |c_{\rho(k)}|^2}.$$

Hence, if we set $x_n = \sum_{k=1}^{3n} \varphi_k$, then

$$\sigma_n(x_n, \mathcal{D}) = \sqrt{2n} = \sqrt{2}\sqrt{n} = \sqrt{2}\sigma_{2n}(x_n, \mathcal{D})$$

and we can use Theorem 4. □

Theorem 9 *Let X be a quasi-Banach space, $\mathcal{D}$ a dictionary of X and $G_n : X \to X$ $(n \in \mathbb{N})$ a sequence of maps such that $\|x - G_n(x)\| \leq C\sigma_n(x, \mathcal{D})$ for all $x \in X$, all $n \in \mathbb{N}$ and a certain constant $C > 0$. Then the following are equivalent claims:*

- (a) *The approximation scheme $(X, \{\Sigma_n(\mathcal{D})\}_{n=0}^{\infty})$ satisfies Shapiro's theorem.*
- (b) *There exists an infinite sequence of elements $\{x_n\}_{n\in\mathbb{N}_0} \subseteq X$ and a constant $K < \infty$ such that*

$$\|x_n - G_n(x_n)\| \leq K\|x_n - G_{2n}(x_n)\| \ \ (n \in \mathbb{N}_0).$$

PROOF The implication $(b) \Rightarrow (a)$ follows from Theorem 4 and the chain of inequalities

$$\sigma_n(x_n, \mathcal{D}) \leq \|x_n - G_n(x_n)\| \leq K\|x_n - G_{2n}(x_n)\| \leq CK\sigma_{2n}(x_n, \mathcal{D})$$

The proof of $(a) \Rightarrow (b)$ follows analogous steps, based on the chain of inequalities:

$$\|x_n - G_n(x_n)\| \leq C\sigma_n(x_n, \mathcal{D}) \leq CL\sigma_{2n}(x_n, \mathcal{D}) \leq CL\|x_n - G_{2n}(x_n)\|,$$

where the constant $L > 1$, the sequence $\mathbb{N}_0$ and the elements $\{x_n\}_{n\in\mathbb{N}_0} \subset X$ verifying $\sigma_n(x_n, \mathcal{D}) \leq L\sigma_{2n}(x_n, \mathcal{D})$ are given by Theorem 4. □

Now we can state and prove the main result of this section.

Theorem 10 *Let $\mathcal{D} = \{\varphi_n\}_{n=1}^{\infty}$ be a normalized unconditional and democratic Schauder basis of the Banach space X. Then for every non-increasing sequence $\{\varepsilon_n\}_{n=0}^{\infty} \in c_0(\mathbb{N})$ there are elements $x \in X$ such that $\sigma_n(x, \mathcal{D}) \neq \mathbf{O}(\varepsilon_n)$.*

PROOF It follows from Theorem 7 that $\mathcal{D}$ is greedy and from Theorem 9, when applied to the maps $G_n(x) = G_n(x, \mathcal{D}, \rho)$ for $x \in X$ and $\rho \in D(x)$, that in order to check that the approximation scheme associated to n-term approximation with respect to $\mathcal{D}$ satisfies Shapiro's theorem we only need to compare the errors $\|x_n - G_n(x_n)\|$ and $\|x_n - G_{2n}(x_n)\|$ for an adequate sequence of elements $x_n \in X$. We set $x_n = \sum_{k=1}^{3n} \varphi_k$. Then $G_n(x_n) = \sum_{k=1}^{n} \varphi_k$ and $G_{2n}(x_n) = \sum_{k=1}^{2n} \varphi_k$, so that

$$\begin{aligned}
\|x_n - G_n(x_n)\| &= \left\| \sum_{k=n+1}^{3n} \varphi_k \right\| \\
&\leq M \left(\left\| \sum_{k=n+1}^{2n} \varphi_k \right\| + \|x_n - G_{2n}(x_n)\| \right) \\
&\leq M \left(C\|x_n - G_{2n}(x_n)\| + \|x_n - G_{2n}(x_n)\| \right) \\
&= K\|x_n - G_{2n}(x_n)\|,
\end{aligned}$$

since $\mathcal{D}$ is democratic. This ends the proof. $\square$

Sometimes it is possible to prove, for a dictionary $\mathcal{D}$ which is not a Schauder basis, that the approximation scheme $(X, \{\Sigma_n(\mathcal{D})\}_{n=0}^{\infty})$ satisfies Shapiro's theorem, but a truly general result is still a (possibly difficult) open question. We include here a case where the result is easy to get although the dictionary is highly redundant.

Theorem 11 *Let $X = L^\infty(0,1)$ and let $\mathcal{D} = \{\chi_I : I = [a,b], 0 \leq a < b \leq 1\}$ be the set of characteristic functions associated to the non-degenerate subintervals of $[0,1]$. Then $(X, \{\Sigma_n(\mathcal{D})\}_{n=0}^{\infty})$ satisfies Shapiro's theorem.*

PROOF They key idea for the proof is to use the strong connection that exists between n-term approximation with the elements of this dictionary and the approximation by splines with free knots. In fact, if $f = \sum_{k=1}^{n} a_k \chi_{I_k}$ is an element of $\Sigma_n(\mathcal{D})$ then f can be decomposed as a superposition of at most $2n+1$ characteristic functions associated to a set of non-degenerate intervals with pairwise disjoint interiors. In particular, this implies that $f \in \mathcal{S}_{4n+2,1}(0,1)$, where $\mathcal{S}_{n,r}(I)$ denotes the set of polynomial splines of degree $< r$ with n free knots on the interval I. The proof of this fact is by induction on n: For $n = 1$ it is obvious. We assume the result for $n = m-1$ and we take $n = m$. If $f = \sum_{k=1}^{m} a_k \chi_{I_k}$ belongs to $\Sigma_m(\mathcal{D})$ then $f = a_m \chi_{I_m} + g$, where $g \in \Sigma_{m-1}(\mathcal{D})$. Clearly, it follows from the induction hypothesis that $g = \sum_{k=1}^{2m-1} b_k \chi_{J_k}$ for certain coefficients $\{b_k\}_{k=1}^{2m-1}$ and non-degenerate intervals with pairwise disjoint interiors, $\{J_k\}_{k=1}^{2m-1}$. Now, the end points of the interval I_m belong, in the worst case, to two distinct intervals J_k. This means that in the worst case we will need to add two more intervals to the representation of f as a superposition of characteristic functions associated to a set of non-degenerate intervals with pairwise disjoint interiors, which proves the claim.

We have already proved that $\Sigma_n(\mathcal{D}) \subseteq \mathcal{S}_{4n+2,1}(0,1)$ so that, to conclude the proof, we only need to prove that the approximation scheme $(L^\infty(0,1), \{\mathcal{S}_{4n+2,1}(0,1)\}_{n=1}^{\infty})$ satisfies Shapiro's theorem. This fact was already mentioned to be true at the very end of section 2 of this paper. We include the proof here just for the sake of completeness.

Let $h(n)$ be a natural number and $f_n(t) = \sin(h(n)\pi t)$. This function equioscillates $h(n)$ times inside the interval $[0,1]$. Moreover, the points of equioscillation of f_n are uniformly distributed on the interval $[0,1]$. On the other hand, if $g \in \mathcal{S}_{8n+4,1}(0,1)$ then there exists at least an interval $I(g) \subset (0,1)$

of length $\geq \frac{1}{8n+4}$ where g is constant. Hence, if we take $h(n)$ big enough then f_n equioscillates as many times as we want on this interval $I(g)$. It follows from the alternation Tchebychev's theorem that if we take $h(n)$ big enough then for every $g \in \mathcal{S}_{8n+4,1}(0,1)$ we have that

$$\|f_n - g\|_{L^\infty(0,1)} \geq \|f_n - g\|_{L^\infty(I(g))} \geq \|f_n\|_{L^\infty(I(g))} = 1 = \|f_n\|_{L^\infty(0,1)}.$$

In particular, this means that $1 = E(f_n, \mathcal{S}_{4n+2,1}(0,1)) = E(f_n, \mathcal{S}_{8n+4,1}(0,1)) = \|f_n\|_{L^\infty(0,1)}$, so that we can use Theorem 4 to claim that $(L^\infty(0,1), \{\mathcal{S}_{4n+2,1}(0,1)\}_{n=1}^\infty)$ satisfies Shapiro's theorem. This ends the proof. $\square$

It is clear that a multidimensional version of Theorem 11 also holds true (it just requires more notation). On the other hand, a version of this theorem for the L^p-norm, with $1 < p < \infty$, is still an open question.

5 A comparison with Brundyi's theorem

Let us prove that Brundyi's condition (1) implies our jump condition $E(x_n, A_n) \leq CE(x_n, A_{K(n)})$ for general approximation schemes (and not just for the case $K(n) = 2n$, which is the only one included in Brundyi's theorem). Indeed, from the use of (1) for $A_{K(n)}$ we know that there exists an element $x_n \in A_{K(n)+1}$ such that $\|x_n\| = 1$ and $E(x_n, A_{K(n)}) \geq \gamma$. Hence, taking $C = 1/\gamma$ we have that, for all $n \in \mathbb{N}$,

$$E(x_n, A_n) \leq 1 = C\gamma \leq CE(x_n, A_{K(n)}),$$

as we wanted to prove. In the opposite direction we have the following result:

Theorem 12 *There exists an approximation scheme that satisfies Shapiro's theorem and does not satisfy Brundyi's condition (1). Moreover, this approximation scheme can be taken verifying $A_n + A_m \subseteq A_{n+m}$ for all n, m.*

Proof We take $X = c_0(\mathbb{N})$ with the usual norm

$$\|(a_n)_{n=0}^\infty\| = \sup_{n\in\mathbb{N}} |a_n|$$

and we introduce the cones B_n given by $B_0 = \{\mathbf{0}\}$, $B_1 = \{(x_1, 0, \cdots, 0, \cdots) : x_1 \in \mathbb{R}\}$ and, for $n \geq 1$,

$$B_{n+1} = \{(x_1, \cdots, x_{n+1}, 0, \cdots) : (x_1, \cdots, x_n) \in \mathbb{R}^n \text{ and } |x_{n+1}| \leq \frac{\sup_{k\leq n} |x_k|}{n+1}\}.$$

Let us also introduce the cones $\Pi_n = \{(x_1, \cdots, x_n, 0, \cdots) : (x_1, \cdots, x_n) \in \mathbb{R}^n\}$. Finally, we consider the approximation scheme $(X, \{A_n\}_{n=0}^\infty)$, where $A_0 = B_0$, $A_1 = B_1 = \Pi_1$, $A_2 = B_2$, $A_3 = \Pi_2$, $A_4 = B_3$, $A_5 = \Pi_3$, $\cdots$.

It is clear that $A_n + A_m \subset A_{\max\{n,m\}+1} \subset A_{n+m}$. Moreover, the chain of inclusions

$$A_0 \subset A_1 \subset A_2 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots c_0(\mathbb{N})$$

is just a new way to write the chain of inclusions

$$B_0 \subset \Pi_1 \subset B_2 \subset \Pi_2 \subset B_3 \subset \Pi_3 \subset \cdots .$$

Furthermore, it is trivial to check that

$$\lim_{n\to\infty} dist(B_{n+1} \cap S(c_0(\mathbb{N})), \Pi_n) = 0,$$

so that

$$\inf_{n\in\mathbb{N}} dist(A_{n+1} \cap S(c_0(\mathbb{N})), A_n) = 0.$$

This means that Brundyi's condition does not hold true for this approximation scheme. On the other hand, it is clear that we can use the classical Bernstein's theorem for the chain of subspaces of X

$$\{0\} \subset \Pi_1 \subset \cdots \subset \Pi_n \subset \Pi_{n+1} \subset \cdots \subset c_0(\mathbb{N})$$

so that the approximation scheme $(X, \{A_n\})$ satisfies Shapiro's theorem. □

6 Approximation schemes that do not satisfy Shapiro's theorem

We have already seen that many classical approximation schemes satisfy Shapiro's theorem. This leads to the following natural question: Do there exists approximation schemes not satisfying Shapiro's theorem? The first known example, as far as we know, is get as a consequence of a famous (and very difficult) result by Pisier. He proved [12, 13] the existence of Banach spaces X with the property that every compact operator $T \in \mathcal{K}(X,X)$ is nuclear. Now, it is well known that the sequence of approximation numbers $\{a_n(T)\}_{n=1}^{\infty}$ of any nuclear operator T belongs to $\ell^1(\mathbb{N})$. It follows that $a_n(T) = \mathbf{O}(1/n)$ since these numbers form a decreasing sequence.

Although this proof is impeccable, it needs to use a very strong result. Thus, it would be nice to have an easy example of approximation scheme that does not satisfy Shapiro's theorem. We solve this question right now.

Example. We take $X = c_0(\mathbb{N})$ with the usual norm and we introduce the cones $A_n = \{(a_k)_{k=0}^{\infty} \in c_0(\mathbb{N}) : \#\{a_k\}_{k=0}^{\infty} \leq n\}$, $n = 1, 2, \cdots$ (for example, the constant sequence $a_k = 1$, $k = 0, 1, 2, \cdots$ belongs to A_1). Obviuously, $(c_0(\mathbb{N}), \{A_n\})$ is an approximation scheme with jump function $K(n) = n^2$. Let $x = (x_n) \in c_0(\mathbb{N})$ and let $n \in \mathbb{N}$ be fixed. Let $M = \sup_{k\in\mathbb{N}} |x_k|$. We take $c_k = M - k\frac{2M}{n}$, $k = 1, 2, \cdots, n-1$. Then every point $\alpha \in [-M, M]$ satisfies $\min_{1\leq k\leq n-1} |\alpha - c_k| \leq \frac{2M}{n}$. Moreover, $x \in c_0(\mathbb{N})$ implies that there exists $N \in \mathbb{N}$ such that $|x_k| < \frac{2M}{n}$ for all $k > N$. With all this information at hand, we can introduce the sequence $a = (a_k)$ given by:

- For all $k > N$, we set $a_k = 0$.
- Let $k \in \{0, 1, 2, \cdots, N\}$. Let $h(k) \in \{1, 2, \cdots, n-1\}$ be such that $|x_k - c_{h(k)}| = \min_{1\leq j\leq n-1} |x_k - c_j|$. Then we set $a_k = c_{h(k)}$.

It is clear that $a = (a_k) \in A_n$ and $\|x - a\| \leq \frac{2M}{n}$. Hence $E(x, A_n) = \mathbf{O}(\frac{1}{n})$, which was our objective.

References

[1] **S. N. Bernstein,** Sur l'ordre de la meilleure approximation des fonctions continues par les polinômes de degré donnè, Mén. Acad. royale Belg., **4** (1912) 1-104.

[2] **S. N. Bernstein,** Sur le probleme inverse de la théorie de la meilleure approximation des functions continues. Comtes Rendus, **206** (1938) 1520-1523.(See also: Ob obratnoi zadache teorii nailuchshego priblizheniya nepreryvnykh funksii, Sochineniya Vol II (1938) 292-294).

[3] **Yu. A. Brudnyi, N. Ya. Krugljak,** *Interpolation functors and interpolation spaces, Vol I* North-Holland Math. Libr. **47** 1991.

[4] **R. Gribonval, M. Nielsen,** Some remarks on nonlinear approximation with Schauder bases, East J. on Approx., **7** No. 2 (2001) 267-285.

[5] **R. Gribonval, M. Nielsen,** Nonlinear approximation with dictionaries. I. Direct estimates, J. of Fourier Anal. and Appl., **10** (1) (2004)

[6] **S. V. Konyagin, V. N. Temlyakov,** A remark on greedy approximation in Banach spaces, East. J. Approx. **5** 3 (1999) 365-379.

[7] **G. Lewicki,** Bernstein's "Lethargy" theorem in metrizable topological linear spaces, Monatshefte fr Mathematik 113 (1992) 213-226.

[8] **G. Lewicki,** A theorem of Bernstein . Iagellon. Acta Math. 27 (1988) 23-27.

[9] **B. Micherda,** Bernstein's "Lethargy" theorems in SF-spaces, Zeitschrift fr Analysis und ihre Anwendungen, 22 (1) (2003) 3-16.

[10] **W. N. Nikolskii,** On some properties of reflexive spaces, Uchen. Zap. Kalinin. Gos. Ped. Inst. **29** (1963) 121-125. (Russian)

[11] **W. N. Nikolskii,** Some remarks on spaces with (B) property, Uchen. Zap. Kalinin. Gos. Ped. Inst. **39** (1964) 48-52. (Russian)

[12] **G. Pisier,** Counterexamples to a conjecture of Grothendiek, Acta Math. **151** (1983) 181-208.

[13] **G. Pisier,** "Factorization of linear operators and Geometry of Banach spaces", Regional Conf. Series in Math. Amer. Math Soc. **60** (1986).

[14] **W. Plesniak,** Quasianalytic functions in the sense of Bernstein, Dissertationes Math. 147 (1977) 1-70.

[15] **H. S. Shapiro,** Some negative theorems of Approximation Theory, Michigan Math. J. **11** 211-217 (1964).

[16] **I. Singer,** "Best approximation in normed linear spaces by elements of linear subspaces", Springer Verlag, New York, 1970.

[17] **V. N. Temlyakov,** The best m-term approximation and greedy algorithms, Advances in Computational Mathematics 8 (1998) 249-265.

[18] **I. S. Tjuriemskih,** B property of Hilbert spaces, ome properties of the Tchebyshev subspaces of a Banach space. Uchen. Zap. Kalinin. Gos. Ped. Inst. **39** (1964) 53-64. (Russian)

J. M. Almira
Departamento de Matemáticas. Universidad de Jaén.
E.P.S. Linares, C/Alfonso X el Sabio, 28
23700 Linares (Jaén) Spain
email: jmalmira@ujaen.es